A Localized Particle Filter for High-Dimensional Nonlinear Systems

JONATHAN POTERJOY

Advanced Study Program, National Center for Atmospheric Research,* Boulder, Colorado

ABSTRACT

This paper presents a new data assimilation approach based on the particle filter (PF) that has potential for nonlinear/non-Gaussian applications in geoscience. Particle filters provide a Monte Carlo approximation of a system's probability density, while making no assumptions regarding the underlying error distribution. The proposed method is similar to the PF in that particles—also referred to as ensemble members—are weighted based on the likelihood of observations in order to approximate posterior probabilities of the system state. The new approach, denoted the local PF, extends the particle weights into vector quantities to reduce the influence of distant observations on the weight calculations via a localization function. While the number of particles required for standard PFs scales exponentially with the dimension of the system, the local PF provides accurate results using relatively few particles. In sensitivity experiments performed with a 40-variable dynamical system, the local PF requires only five particles to prevent filter divergence for both dense and sparse observation networks. Comparisons of the local PF and ensemble Kalman filters (EnKFs) reveal advantages of the new method in situations resembling geophysical data assimilation applications. In particular, the new filter demonstrates substantial benefits over EnKFs when observation networks consist of densely spaced measurements that relate nonlinearly to the model state—analogous to remotely sensed data used frequently in weather analyses.

1. Introduction

Ensemble filters and smoothers provide a means of estimating the probability density of a system state, given observations and a numerical model for the dynamical system. These methods are used frequently for data assimilation in geophysical systems, such as the earth's atmosphere and ocean; examples include the ensemble Kalman filter (EnKF; Evensen 1994; Houtekamer and Mitchell 1998; Evensen and van Leeuwen 2000) and ensemble-variational hybrid schemes (e.g., Hamill and Snyder 2000; Lorenc 2003; Buehner 2005; Liu et al. 2008; Zhang et al. 2009). The above-mentioned strategies approximate error distributions for observations and model forecasts using Gaussian probabilities, causing them to be suboptimal when the model dynamics are nonlinear, or

when the assimilated observations relate nonlinearly to the model state.

Despite their limitations, techniques that rely on Gaussian assumptions have performed well for operational weather prediction and research (e.g., Buehner et al. 2010; Bishop and Hodyss 2011; Clayton et al. 2013; Kuhl et al. 2013; Wang and Lei 2014). Nevertheless, it is uncertain whether these methods are the best means of assimilating observations as computational resources allow for increasingly large ensembles. For example, Miyoshi et al. (2014) show that ensembles of ~1000 members can provide accurate representations of non-Gaussian prior probabilities that occur for atmospheric quantities such as moisture. Ensemble statistics can also exhibit large deviations from Gaussianity when projected into observation space by nonlinear operators, as demonstrated by Pires et al. (2010). This problem limits the effectiveness of Gaussian filters for assimilating remotely sensed data such as satellite radiances and radar reflectivity.

One filtering approach used frequently for low-dimensional systems is the particle filter (PF) [see Doucet et al. (2001) and van Leeuwen (2009) for a review]. The PF provides posterior weights to ensemble

Corresponding author address: Dr. Jonathan Poterjoy, NCAR, P.O. Box 3000, Boulder, CO 80307.
E-mail: poterjoy@ucar.edu

members (also denoted particles), which reflect the likelihood of observations given each member. These weights provide a means of estimating properties of the posterior error distribution with no assumptions regarding error distribution for the prior model state. The weights also determine which members to remove or duplicate in PF algorithms that include resampling steps, with the goal of retaining ensemble members in regions of high probability only.

Despite the many attractive properties of PFs, they require an ensemble size that increases exponentially with the dimension of the system (Snyder et al. 2008, hereafter S08). Naively using too small of an ensemble will cause the weights to collapse on to a single particle, rendering the posterior representation meaningless. The failure of the PF as an affordable filter for high-dimensional systems remains a fundamental obstacle for applying this method on problems in geophysics, such as data assimilation for operational weather prediction models.

Several strategies for overcoming the dimensionality challenges of PFs have been proposed recently. A common approach is to prevent the weight collapse by manipulating the transition density between data assimilation cycles (i.e., through the propagation of particles in time), or by drawing particles from a proposal density conditioned on the current observations. The equal-weights PF (van Leeuwen 2010) and the implicit PF (Chorin et al. 2010) are both examples of this approach. Other methods use a PF for filtering strongly non-Gaussian portions of the state space, while maintaining an EnKF for high-dimensional quasi-Gaussian quantities. Filters based on this idea include the blended PF (Majda et al. 2014) and the hybrid particle-ensemble Kalman filter (Slivinski et al. 2015). Frei and Künsch (2013) also propose a filter that combines aspects of particle filtering and Kalman filtering theory. Their method transitions between an EnKF and a PF, where the EnKF component is weighted more when filter degeneracy is likely to occur. Another strategy, introduced in Reich (2013), is to avoid the random sampling aspect of PFs by solving an optimal transportation problem for transforming prior particles into posterior particles. While Ades and van Leeuwen (2015) have shown some success applying the equal-weights PF for high-dimensional systems, filters based on the PF framework have yet to be proven practical for real geoscience applications.

Strategies used to develop EnKFs into effective tools for high-dimensional data assimilation may provide further insight into how to overcome obstacles for PFs. Both the EnKF and PF use model realizations of the system state to estimate prior errors, except the EnKF relies solely on the ensemble mean and covariance to approximate a probability density function (pdf). The

success of EnKFs in high-dimensional systems is due in part to the use of covariance localization (Houtekamer and Mitchell 2001; Hamill and Whitaker 2001). In this context, localization refers to the tapering of ensemble-estimated covariances as a function of distance, which is particularly beneficial for applications containing a large spatial dimension. By exploiting the fact that the signal-to-noise ratio of covariances[1] tends to decrease at large distances, localization reduces the influence of observations on distant state variables during the posterior update. A typical approach is to represent prior covariances using an element-wise product of ensemble covariances and a correlation function with compact support (Gaspari and Cohn 1999). Particle filters do not rely explicitly on prior covariances, so localization in the same manner is not feasible. Nevertheless, a similar strategy for reducing the dimensionality constraints of this filter may be required before it can become a practical data assimilation method for high-dimensional problems. Localization strategies have already been adopted in several non-Gaussian filters, including the local–local Gaussian mixture filter (Bengtsson et al. 2003), the rank histogram filter (Anderson 2010), and the moment-matching filter (Lei and Bickel 2011).

The current study presents a new filtering approach based on the PF that has potential for data assimilation applications encountered frequently in geoscience. The proposed method (denoted local PF) provides a general Bayesian update of particles in regions near the physical location of observations, while preserving prior particles away from observations. In doing so, the filter achieves localized updates in a manner similar to covariance localization in EnKFs. The local PF is designed to be both effective for reasonable ensemble sizes and computationally inexpensive, thus making it a viable technique for large geophysical models. Furthermore, the new filter is developed using the National Center for Atmospheric Research (NCAR) Data Assimilation Research Testbed (DART) software package, which allows for its direct comparison with other ensemble filters. This manuscript introduces the local PF algorithm and presents cycling data assimilation experiments performed using the 40-variable Lorenz (1996) model (denoted L96). The performance benefits of the new filter are assessed using the ensemble adjustment Kalman filter (EAKF; Anderson 2001) as a benchmark.

The manuscript is organized in the following manner. Section 2 presents the PF in its most basic form and describes localization in the context of the dimensionality

---

[1] The signal-to-noise ratio is the true covariance divided by the variance of its sample estimate.

problem that places practical constraints on this approach. Section 3 introduces the local PF update algorithm, which incorporates the localization strategy discussed in section 2. Section 4 provides results from cycling data assimilation experiments and compares the local PF with the EAKF over a range of configurations. The last section summarizes results from this study and discusses the potential of the new filter for large applications.

2. Particle filtering for high-dimensional systems

a. The particle filter

This section introduces the framework for the most basic sequential PF. Let $\mathbf{x}$ be a random vector of length N_x representing all prognostic state variables of a system at a given time, and let $\mathbf{y}$ be a vector of length N_y containing observations. The observations are related to the true state $\mathbf{x}^t$ through

$$\mathbf{y} = H[\mathbf{x}^t] + \epsilon, \tag{1}$$

where H is a measurement operator that maps the state vector into observation space and ϵ is the measurement error. The only information needed to estimate the probability density of $\mathbf{x}$ conditioned on new observations is a prior probability $p(\mathbf{x})$ and the likelihood of the observations $p(\mathbf{y}|\mathbf{x})$. Bayes's theorem provides an expression for the posterior probability of $\mathbf{x}$ given all information up to the current time:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{\int p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\,d\mathbf{x}}. \tag{2}$$

One means of obtaining $p(\mathbf{x}|\mathbf{y})$ is through a Monte Carlo approximation of the distributions in (2). For example, provided with N_e particles sampled from $p(\mathbf{x})$ (denoted $\mathbf{x}_n, n = 1, \ldots, N_e$), $p(\mathbf{x})$ can be constructed as a sum of delta functions centered on each particle:

$$p(\mathbf{x}) \approx \frac{1}{N_e} \sum_{n=1}^{N_e} \delta(\mathbf{x} - \mathbf{x}_n). \tag{3}$$

Likewise, the posterior density can be approximated using

$$p(\mathbf{x}|\mathbf{y}) \approx \sum_{n=1}^{N_e} \frac{w_n}{W} \delta(\mathbf{x} - \mathbf{x}_n), \tag{4}$$

with normalized weights, w_n/W, provided by

$$w_n = p(\mathbf{y}|\mathbf{x}_n), \quad \text{and} \tag{5}$$

$$W = \sum_{m=1}^{N_e} w_m. \tag{6}$$

The normalization of weights by (6) serves as an estimator for the denominator of (2). The particle approximation of $p(\mathbf{x}|\mathbf{y})$ using (4)–(6) follows the simplest form of sequential importance sampling in which the proposal density is chosen as the prior probability density, and the weights from the previous filtering time are assumed to be equal (Doucet et al. 2001). The assumption of equal prior weights is satisfied if new particles are resampled from the posterior each filtering time—as will be the case for the filtering method proposed in this study.

The particle weights can also be applied to estimate moments of the posterior error distribution:

$$\overline{f(\mathbf{x})} = \int f(\mathbf{x})p(\mathbf{x}|\mathbf{y})\,d\mathbf{x},$$

$$\approx \sum_{n=1}^{N_e} \frac{w_n}{W} f(\mathbf{x}_n). \tag{7}$$

For example, (7) approximates the posterior mean using a weighted sum of particles: $\overline{\mathbf{x}}^a \approx \sum_{n=1}^{N_e}(w_n/W)\mathbf{x}_n$. Functions of this type will be referred to as "moment-estimating functions" in this manuscript.

b. Filter degeneracy in particle filters

The filter described in section 2a is completely general in that it makes no assumptions regarding the error distributions needed to estimate and sample from the posterior density. Nevertheless, the N_e required to prevent the collapse of the weights onto a single particle makes the PF impractical for high-dimensional systems. The convergence of the largest posterior weight to unity will be referred to as "filter degeneracy" in this discussion.

Under weak assumptions, Bengtsson et al. (2008) prove that filter degeneracy occurs when N_e increases subexponentially with the state dimension. Their proof holds for all forms of observation likelihoods, including cases when observation error distributions are specified to have heavy tails to delay the weight collapse (van Leeuwen 2003). Bickel et al. (2008) extend the formal proof in Bengtsson et al. (2008) to include more general conditions. They also emphasize that the N_e required to prevent filter degeneracy depends on the sum of singular values of the observation-space prior covariance, rather than the state dimension alone. S08 provide further asymptotic analysis of the PF for independent and identically distributed observation likelihoods, and for cases when both the prior and observation errors are Gaussian. Their study provides more specific criteria for avoiding the weight collapse, which requires an exponential increase of N_e with the variance of the observation log-likelihood (denoted τ^2). Here, τ^2 can be thought of as an estimate of the state dimension depicted by measurements in $\mathbf{y}$.

c. Localization as a means of preventing filter degeneracy

As discussed in section 1, EnKFs rely on covariance localization to stabilize the filter when N_e is small relative to the state dimension. S08 speculate that a similar strategy may be needed in order to reduce the dimensionality constraint on PFs. To describe how localization may be achieved in a PF framework, consider the case where a single observation, y, is available and posterior weights are calculated using the likelihood of the observation given each particle. One means of achieving localization is to extend the original weights from scalars to vectors of length N_x, which will be denoted by $\boldsymbol{\omega}_n$ in this manuscript. The resulting vectors form the columns of a $N_x \times N_e$ weighting matrix, and are constructed to reflect the local influence of observations on the posterior estimate. As in EnKFs, the influence of observations on neighboring state-space updates is specified using prior knowledge of the "physics" of the system (e.g., physical length scales contributing to spatial correlations in the prior). This form of localization is achieved by including the function, $l[y, x_j, r]$, in the calculation of the jth elements of each $\boldsymbol{\omega}_n$ and their normalization vector $\boldsymbol{\Omega}$:

$$\omega_{n,j} = [p(y \,|\, x_{n,j}) - 1]l[y, x_j, r] + 1, \quad \text{and} \tag{8}$$

$$\Omega_j = \sum_{m=1}^{N_e} \omega_{m,j}. \tag{9}$$

The localization function has a maximum value of 1 when the Euclidean distance between y and x_j is 0, and decays to 0 when y and x_j are far apart; the rate of this decay is controlled by the parameter r. In practice, $l[y, x_j, r]$ should be a smooth function with compact support. The current study uses (4.10) of Gaspari and Cohn (1999) for $l[y, x_j, r]$, which has a Gaussian-type structure with a width specified by r.

The equation chosen for forming the vector weights is motivated by two factors. The first advantage of (8) is that it accomplishes the original goal of localizing information spatially. Using the symbol ./ to denote element-wise division, the normalized weights $\boldsymbol{\omega}_n./\boldsymbol{\Omega}$ reflect the observation likelihood near y and the prior weights $(1/N_e)$ away from y. The second motivating factor involves the computation of weights when given multiple observations. Assuming observation errors are independent, $p(\mathbf{y} \,|\, \mathbf{x}_n)$ can be written $\prod_{i=1}^{N_y} p(y_i \,|\, \mathbf{x}_n)$, where y_i is the ith observation in $\mathbf{y}$. The values for the jth elements of the weights given the ith observation are then found sequentially by

$$\omega_{n,j}^{(y_i)} = \prod_{q=1}^{i} \{[p(y_q \,|\, x_{n,j}^{(y_0)}) - 1]l[y_q, x_j, r] + 1\},$$

$$= \omega_{n,j}^{(y_{i-1})} \{[p(y_q \,|\, x_{n,j}^{(y_0)}) - 1]l[y_q, x_j, r] + 1\}, \tag{10}$$

$$\Omega_j^{(y_i)} = \sum_{n=1}^{N_e} \omega_{n,j}^{(y_i)}, \tag{11}$$

where superscript (y_i) refers to quantities that reflect all observations up to y_i and $\mathbf{x}_n^{(y_0)}$ is the prior ensemble before assimilating any observations in $\mathbf{y}$. For applications where many observations are assimilated over a large spatial domain, most values in the product of (10) will be equal to 1. The resulting weight equation is numerically stable for large N_y, because the rate at which this product approaches zero depends only on the number of observations within the localization region defined by $l[y_q, x_i, r]$. After applying (10) and (11) to calculate the weights, posterior quantities are approximated using

$$\overline{f(\mathbf{x})} \approx \sum_{n=1}^{N_e} (\boldsymbol{\omega}_n./\boldsymbol{\Omega}) \circ f(\mathbf{x}_n), \tag{12}$$

where $\circ$ represents an element-wise vector product. The weighting vectors provide information regarding the marginal probabilities for each state variable only, so (12) cannot estimate multivariate properties, such as covariance. Because of this shortcoming, the local PF algorithm outlined in section 3 requires a bootstrap resampling step to make multivariate corrections to particles.

To illustrate how the localized weight equation affects the PF estimate of posterior quantities, consider the case where a single observation is assimilated to estimate the posterior mean and variance of a random vector $\mathbf{x}$ of size 100. Let the observation, $y = 1$, provide an estimate of state variable 50 with an error variance equal to unity. Also, let the true prior pdf be given by $p(\mathbf{x}) = N(\boldsymbol{\mu}, \mathbf{B})$, where $\boldsymbol{\mu}$ is a 100-element vector of zeros and $\mathbf{B}$ is a covariance matrix containing diagonal elements equal to unity. For this demonstration, off-diagonal elements in $\mathbf{B}$ are modeled using a product of Gaussian and sinusoidal functions (see Fig. 1). The resulting covariance matrix is sparse and contains arbitrary structure near the diagonal elements, which is typical for most geophysical applications. Using 50 and 1000 samples drawn from the prior, red lines in Figs. 2a–d show the posterior mean and variance estimated using the standard PF approach given by (5)–(7); the optimal least squares solution using the true mean and covariance is plotted in black for reference. The spatial structure of posterior quantities is influenced largely by
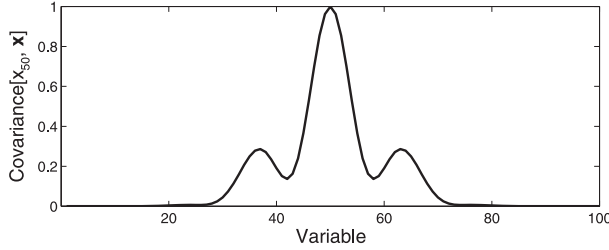
FIG. 1. Correlations between state variable 50 and all other variables in **x** modeled using the product of Gaussian and sinusoidal functions.

the correlations specified in **B** (see black line in Fig. 1), so the optimal solution is equal to the prior where variables are uncorrelated with the observed variable. The small ensemble provides relatively accurate results near the observation, but overestimates the impact of the observation on the posterior far from state variable 50. Even with 1000 particles, the PF mean and variance are relatively noisy in regions where the observation should have no effect on the posterior. Figures 2e and 2f show posterior quantities estimated using the localizing approach with 50 particles. This example demonstrates that the approximation in (12) preserves the PF solution near the observation, while removing most of the distant errors (Figs. 2e,f). As a result, the local PF captures major components of the update using a much smaller ensemble than would be required otherwise, thus providing a good approximation when large ensembles are not available.

Before a PF method based on localized weights can be conceived, localization must first be shown to remove the filter's exponential dependence on N_e. Section 3 of S08 contains a simple example that illustrates the relationship between N_e and state dimension, which will be reproduced here. The filtering problem of interest uses $N_y = N_x$, for N_x ranging from 10 to 90, and $\mathbf{x}^t \sim N(0, \mathbf{I})$. Using particles and ϵ drawn randomly from $N(0, \mathbf{I})$, the PF is applied with an increasingly large N_e until the posterior mean produces a domain-averaged root-mean-square error (RMSE) smaller than either the prior mean or observations. Following S08, the experiments are performed over 400 trials using $N_e = 10 \times 2^k$, where k is increased until meeting the stopping criteria. These simulations estimate the minimum N_e required for the PF to provide a result that is more useful than the prior and observed information as N_x and N_y increase. Results of these simulations are summarized in Fig. 3 by indicating the mean N_e for each set of trials over various N_x and N_y. Because the observed state variables are independent in this example, these simulations provide an effective demonstration of the PF's behavior for problems with increasingly higher degrees of freedom. As a result, the "no localization" case (black markers in Fig. 3) is a direct replication of Fig. 2 in S08, which demonstrates the exponential increase in N_e required for an increase in system dimension. Localization effectively removes this exponential relationship, as indicated by trials using (10) and (11) to calculate the weights (colored markers in Fig. 3). Experiments are
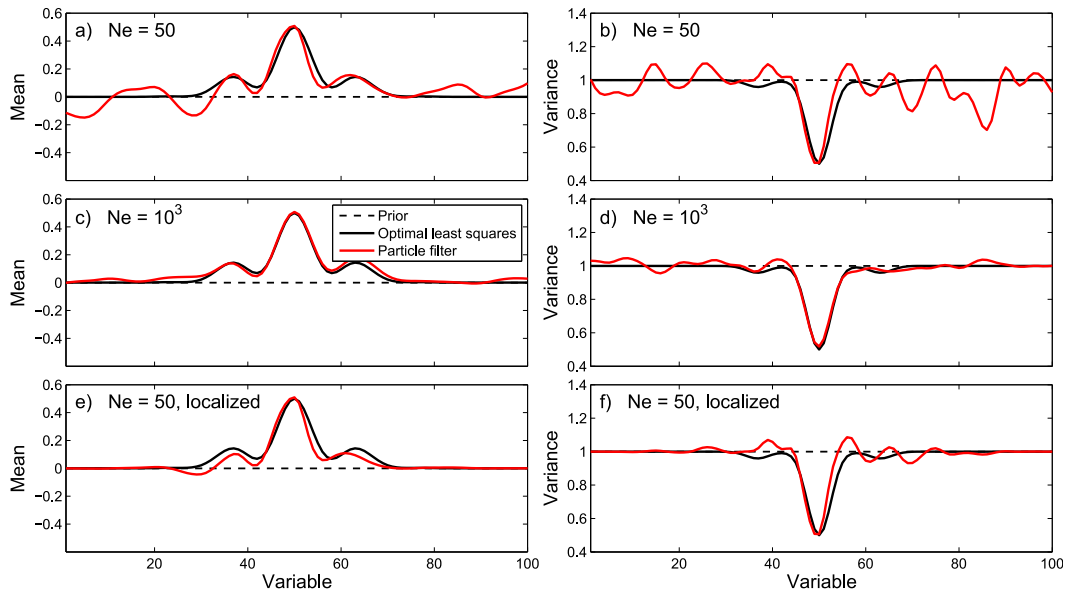
FIG. 2. Mean and variance of the prior solution (dashed black), optimal least squares solution (solid black), and PF solution (red). The PF uses (a),(b) $N_e = 50$; (c),(d) $N_e = 1000$; and (e),(f) $N_e = 50$ with localization.
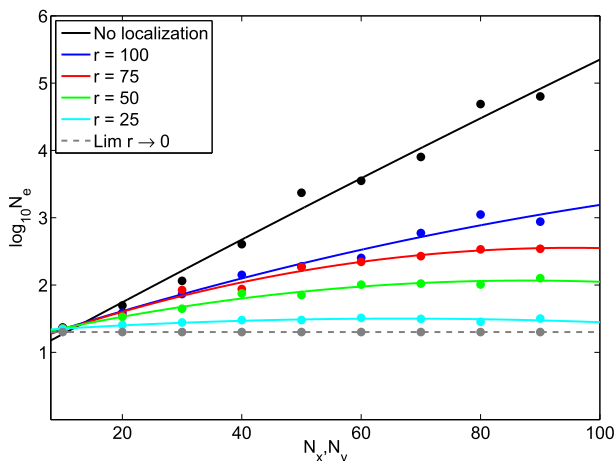
FIG. 3. Markers indicate the N_e required for the PF to produce errors smaller than the prior or observations for a given N_x and N_y. Values are plotted for several different localization radii r, along with a second-degree polynomial fit to the markers.

performed for multiple localization radii r, ranging from near zero to 100, to demonstrate the effects of allowing varying numbers of observations (degrees of freedom) in the update of each state variable. For this application, decreasing r leads to a smaller increase in the N_e required for the posterior to provide accurate results.[2] The optimal localization emerges as the case in which $r \to 0$, because the true prior correlations between neighboring state variables are equal to zero. This case is equivalent to applying the PF independently for every element of $\mathbf{x}$, thus causing the N_e curve to be flat.

In the context of S08's criteria for weight collapse, the extension of weights from scalar to vector quantities means a τ^2 exists for each state variable. Localization limits the impact of distant independent observations on the calculation of weights; therefore, τ^2 at a given grid point depends only on observations within the neighborhood defined by the localization function. Provided that most observations lie outside each neighborhood, the N_e needed for large N_x and N_y can be reduced substantially.

3. The local particle filter

The localized weight equations introduced in section 2c provide a means of estimating posterior quantities using small ensembles for high-dimensional systems that have finite prior correlation length scales between

---

[2] Some data points in Fig. 3 show a slight decrease in N_e between an increasing set of dimensions; this result is caused entirely by sampling error.

spatially separated variables. Generating equally likely samples from the posterior density, however, presents another challenge for PFs. A typical sampling strategy for low-dimensional stochastic systems is to remove particles with small weights and duplicate particles with large weights; the simplest example is the bootstrap filter (Gordon et al. 1993). A similar approach is applied for the local PF, except localization adds complexity to the process because a unique weight exists for each element of the state vector. Therefore, particles must be modified to fit characteristics of the posterior given by (12). The approach taken in this study is to process observations at each filter time serially, while recursively updating the particles. This strategy follows two steps: 1) apply bootstrap resampling for each observation and merge prior particles with resampled particles to generate samples from a distribution with the approximate first- and second-order moments; and 2) use probability mapping to adjust the new particles so that they are consistent with the marginal probabilities given by the set of posterior weights for each variable. The first step is similar to several previously proposed methods, which form samples to approximate the first two moments of the posterior pdf (e.g., Xiong et al. 2006; Nakano et al. 2007; Lei and Bickel 2011). One additional objective of the first step is to preserve the sampled particles near each observation, so that the updated particles approach the bootstrap filter solution near observations. The second step provides higher-order corrections to the particles not considered during the first step.

a. Sampling and merging step

To describe the first part of the algorithm, consider the adjustment of particles associated with the ith observation. The prior error distribution before assimilating y_i is approximated with N_e equally likely particles that represent samples from the probability density given all observations up to y_{i-1}; these particles will be denoted by $\mathbf{x}_n^{(y_{i-1})}$ for $n = 1, \ldots, N_e$. To maintain consistency with the localized weighting vectors, the local PF must create posterior particles that satisfy the Bayesian solution in regions of the state space assumed to be influenced by y_i. Likewise, regions of the state space assumed to be independent of y_i need to maintain characteristics of the prior. To achieve this result, a scalar weight $\tilde{w}_n = p(y_i | \mathbf{x}_n^{(y_{i-1})})$ is first calculated for each particle, then normalized by $\tilde{W} = \sum_{n=1}^{N_e} \tilde{w}_n$. These weights are then used to sample N_e particles with replacement to provide posterior particles that would result from applying the bootstrap filter. Updates are then made to the prior particles in a manner that is consistent with the bootstrap filter solution near observations, and

the first two moments of the localized posterior solution in the neighborhood of the observation:

$$\mathbf{x}_n^{(y_i)} = \overline{\mathbf{x}}^{(y_i)} + \mathbf{r}_1 \circ (\mathbf{x}_{k_n}^{(y_{i-1})} - \overline{\mathbf{x}}^{(y_i)}) + \mathbf{r}_2 \circ (\mathbf{x}_n^{(y_{i-1})} - \overline{\mathbf{x}}^{(y_i)}), \quad (13)$$

where, $\overline{\mathbf{x}}^{(y_i)}$ is the posterior mean calculated using (12) and k_n is the index of the nth sampled particle.[3] The new particles are formed as linear combinations of the sampled particles and prior particles using the coefficient vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ of length N_x to specify the influence of localization on the updates. The form chosen for (13) provides a straightforward means of deriving an update equation that satisfies the bootstrap filter solution at the location of observations, and the posterior mean and variance calculated from (12) within the localization region. A solution that satisfies this criteria (see the appendix) is given by the set of equations for the jth elements of $\mathbf{r}_1$ and $\mathbf{r}_2$:

$$r_{1,j} = \sqrt{\dfrac{\sigma_j^{(y_i)^2}}{\dfrac{1}{N_e - 1}\sum_{n=1}^{N_e}[x_{k_n j}^{(y_{i-1})} - \overline{x}_j^{(y_i)} + c_j(x_{n,j}^{(y_{i-1})} - \overline{x}_j^{(y_i)})]^2}},$$
$$(14)$$

$$r_{2,j} = c_j r_{1,j}, \quad \text{and} \quad (15)$$

$$c_j = \dfrac{N_e(1 - l[x_j, y_i, r])}{l[x_j, y_i, r]\tilde{W}}, \quad (16)$$

where $\sigma_j^{(y_i)2}$ is the error variance conditioned on all observations up to y_i. Posterior correlations between state variables are not considered during this formulation, but are provided implicitly through the sampling step of the algorithm. To interpret (13)–(16), consider the asymptotic behavior of $r_{1,j}$ and $r_{2,j}$ as $l[x_j, y_i, r]$ approaches 1 and 0. As $l[x_j, y_i, r] \rightarrow 1$, $c_j \rightarrow 0$, and

$$\lim_{c_j \rightarrow 0} r_{1,j} = \sqrt{\dfrac{\sigma_j^{(y_i)^2}}{\dfrac{1}{N_e - 1}\sum_{n=1}^{N_e}(x_{k_n j}^{(y_{i-1})} - \overline{x}_j^{(y_i)})^2}}, \quad (17)$$

$$\approx 1,$$

because the posterior variance is approximately equal to the sampled particle variance when $l[x_j, y_i, r] = 1$. Likewise, $\lim_{c_j \rightarrow 0} r_{2,j} = 0$, which leads to (13) placing all weight onto the sampled particles. As $l[x_j, y_i, r] \rightarrow 0$, $c_j \rightarrow \infty$, and

$$\lim_{c_j \rightarrow \infty} r_{2,j} = \sqrt{\dfrac{\sigma_j^{(y_i)^2}}{\sqrt{\dfrac{1}{N_e - 1}\sum_{n=1}^{N_e}(x_{n,j}^{(y_{i-1})} - \overline{x}_j^{(y_i)})^2}}}, \quad (18)$$

$$= 1,$$

because the posterior variance is equal to the prior variance when $l[x_j, y_i, r] = 0$. At the same time, $\lim_{c_j \rightarrow \infty} r_{1,j} = 0$, which leads to (13) placing all weight onto the prior particles.

The sampling step provides a means of adjusting particles to fit the general Bayesian posterior solution near the observations. Because each sampled particle is combined with a prior particle, the resulting posterior ensemble contains N_e unique model states, which avoids the collapse of the ensemble variance during the serial assimilation of observations. Random sampling errors introduced during each update step may accumulate after processing several observations. To reduce these errors, the mean and variance terms in (14) are estimated using (10)–(12), which are independent of the sampling and update procedures described above. This part of the algorithm requires storing the prior particles before assimilating the first observation [i.e., $\mathbf{x}_n^{(y_0)}$, $n = 1, \ldots, N_e$] and updating the weighting matrix sequentially with each new observation according to (10) and (11). Last, the index of sampled particles replaces the index of removed particles in (13), so that k_n is equal to n for the first occurrence of each particle selected during sampling. This step ensures that particles that survive the sampling step undergo minimal adjustment by (13).

In addition to localization, the filter's stability for small ensemble sizes can be improved by multiplying $l[x_j, y_i, r]$ by a scalar α, where $\alpha < 1$. This step forces the weights to be more uniform in regions where observations have a large impact on the filter update. The modification to $l[x_j, y_i, r]$ reduces the update of state variables near observations in a manner similar to "relaxation" approaches used in some formulations of the EnKF (Zhang et al. 2004; Whitaker and Hamill 2012). In addition, $\tilde{w}_n$ is replaced with $[p(y_i | \mathbf{x}_n^{(y_{i-1})}) - 1]\alpha + 1$ to maintain consistency between the scalar weights used to resample particles and the vector weights used in the moment-estimating equations. This step also places a minimum on the weights calculated before normalization and increases the number of unique particles selected during the sampling step. Problems arise when errors differ between observations, in which case, the likelihood values can be orders of magnitude different between various observation types. In this case, the weights must be normalized before applying α.

---

[3] Multiple copies of particles may result from the sampling, causing duplicate k_n indices to exist.

b. Probability mapping step

After updating the particles within the localization region using (13), higher-order corrections are then made using probability mapping methods used frequently for debiasing model output (e.g., Ines and Hansen 2006; Piani et al. 2010; Haerter et al. 2011). The method chosen for this study is the kernel density distribution mapping (KDDM) approach developed by McGinnis et al. (2015) for non-Gaussian densities. KDDM operates by mapping a prior sample into a posterior sample that matches the quantiles of a specified posterior distribution—here the desired posterior distribution is defined by the prior particles and their posterior weights. One advantage of KDDM is that when applied separately for each state variable in $\mathbf{x}$, the resulting posterior ensemble contains approximately the same correlations as the prior ensemble.[4] Therefore, univariate KDDM steps can be applied to the particles while maintaining the cross-variable correlations that resulted from the sampling part of the update algorithm described above. For simplicity in notation, denote the jth values of input (prior) and output (posterior) particles as $x_{n,j}^f$ and $x_{n,j}^a$, respectively. Starting from the recently updated particles and weights, KDDM uses the following steps to perform the mapping:

1) Approximate the prior and posterior densities using linear combinations of Gaussian kernels. This step uses a sum of kernels centered on each $x_{n,j}^a$, which are weighted by $1/N_e$ to form a prior pdf (pdf^f) and $\omega_{n,j}^{(y_i)}/\Omega_j^{(y_i)}$ to form a posterior pdf (pdf^a). A fixed kernel bandwidth of 1 is chosen for this study, but different choices may be necessary for more complex filtering problems.

2) Integrate the two pdfs numerically via the trapezoid rule to form the prior cdf (cdf^f) and posterior cdf (cdf^a).

3) Apply cubic spline interpolation to find the prior cdf values at the location of each prior member: $c_{n,j}^f = \text{cdf}^f(x_{n,j}^f)$.

4) Estimate posterior particles by applying cubic spline interpolation to find the inverse of the posterior cdf at each $c_{n,j}^f$: $x_{n,j}^a = \text{cdf}^{a-1}(c_{n,j}^f)$.

The impact of probability mapping on the performance of the filter is discussed in section 4.

c. Algorithm summary and example

Algorithm 1 provides a pseudocode description of the local PF. The major steps of this algorithm are illustrated by the schematic in Fig. 4 for a case where two observations (y_1 and y_2) are assimilated for the 40-variable L96 model using four particles. The filter starts with an ensemble of equally weighted prior particles [denoted $\mathbf{x}_n^{(y_0)}$], shown in model space and observation space in Figs. 4a and 4b, respectively. Particles are first sampled with replacement to select $\mathbf{x}_{k_n}^{(y_0)}$ for $n = 1, \ldots, 4$ (Fig. 4c) based on scalar weights proportional to $p(y_1 \mid \mathbf{x}_n^{(y_0)})$ (Fig. 4d). The next step applies (13) to merge each $\mathbf{x}_n^{(y_0)}$, $\mathbf{x}_{k_n}^{(y_0)}$ pair to get $\mathbf{x}_n^{(y_1)}$ for $n = 1, \ldots, 4$ (Fig. 4e.). The vector coefficients $\mathbf{r}_1$ and $\mathbf{r}_2$ needed for the merge step depend on spatially smooth vectors of posterior weights, which are calculated from (10) and (11) and plotted for each particle in Fig. 4f. By construction, these weights are equal to the standard nonlocalized PF weights near y_1 and $\frac{1}{4}$ away from y_1. After processing the first observation, y_2 is assimilated through the same process, using $\mathbf{x}_n^{(y_1)}$ for $n = 1, \ldots, 4$ as the new prior particles (Figs. 4g–l). Note that the weighting vectors plotted in Fig. 4l for assimilating y_2 also depend on the previous weights in Fig. 4f from (10). These weights are used with the original set of prior particles (from Fig. 4a) to estimate posterior quantities needed for (13). The particles then undergo a final adjustment in Fig. 4m using probability mapping.

Algorithm 1 Localized particle filter algorithm

Input: Initial weighting matrix ($\omega_{n,j} = 1$, $n = 1, \ldots, N_e$, $j = 1, \ldots, N_x$), prior ensemble ($\mathbf{x}_n^{(y_0)}$, $n = 1, \ldots, N_e$), observations $\mathbf{y}$, inflation parameter α, and localization length scale r.

$\quad$ **for** $i = 1:N_y$ **do**

$\quad\quad$ **for** $n = 1:N_e$ **do**

$\quad\quad\quad$ $\tilde{w}_n \leftarrow [p(y_i \mid \mathbf{x}_n^{(y_{i-1})}) - 1]\alpha + 1$

$\quad\quad$ **end for**

$\quad\quad$ $\tilde{W} \leftarrow \sum_{n=1}^{N_e} \tilde{w}^n$

$\quad$ Draw particles $\mathbf{x}_{k_n}^{(y_{i-1})}$, for $n = 1, \ldots, N_e$, from current prior according to weights $\tilde{w}_n/\tilde{W}$

$\quad\quad$ **for** $j = 1:N_x$ **do**

$\quad\quad\quad$ **for** $n = 1:N_e$ **do**

$\quad\quad\quad\quad$ $\omega_{n,j}^{(y_i)} \leftarrow \omega_{n,j}^{(y_{i-1})}\{[p(y_i \mid \mathbf{x}_n^{(y_0)}) - 1]l[y_i, x_j, r]\alpha + 1\}$

$\quad\quad\quad$ **end for**

$\quad\quad\quad$ $\Omega_j^{(y_i)} \leftarrow \sum_{n=1}^{N_e} \omega_{n,j}^{(y_i)}$

$\quad\quad\quad$ $\bar{x}_j^{(y_i)} \leftarrow \sum_{n=1}^{N_e} \frac{\omega_{n,j}^{(y_i)}}{\Omega_j^{(y_i)}} x_{n,j}^{(y_0)}$

$\quad\quad\quad$ $\sigma_j^{(y_i)2} \leftarrow \sum_{n=1}^{N_e} \frac{\omega_{n,j}^{(y_i)}}{\Omega_j^{(y_i)}}[x_{n,j}^{(y_0)} - \bar{x}_j^{(y_i)}]^2$

$\quad\quad\quad$ $c_j \leftarrow \frac{N_e(1 - l[x_j, y_i, r]\alpha)}{l[x_j, y_i, r]\alpha\tilde{W}}$

---

[4] To achieve this result, it is also necessary to center and scale the sample to have a mean of 0 and variance of 1, then recenter and scale the sample after the mapping to reflect the posterior mean and variance.
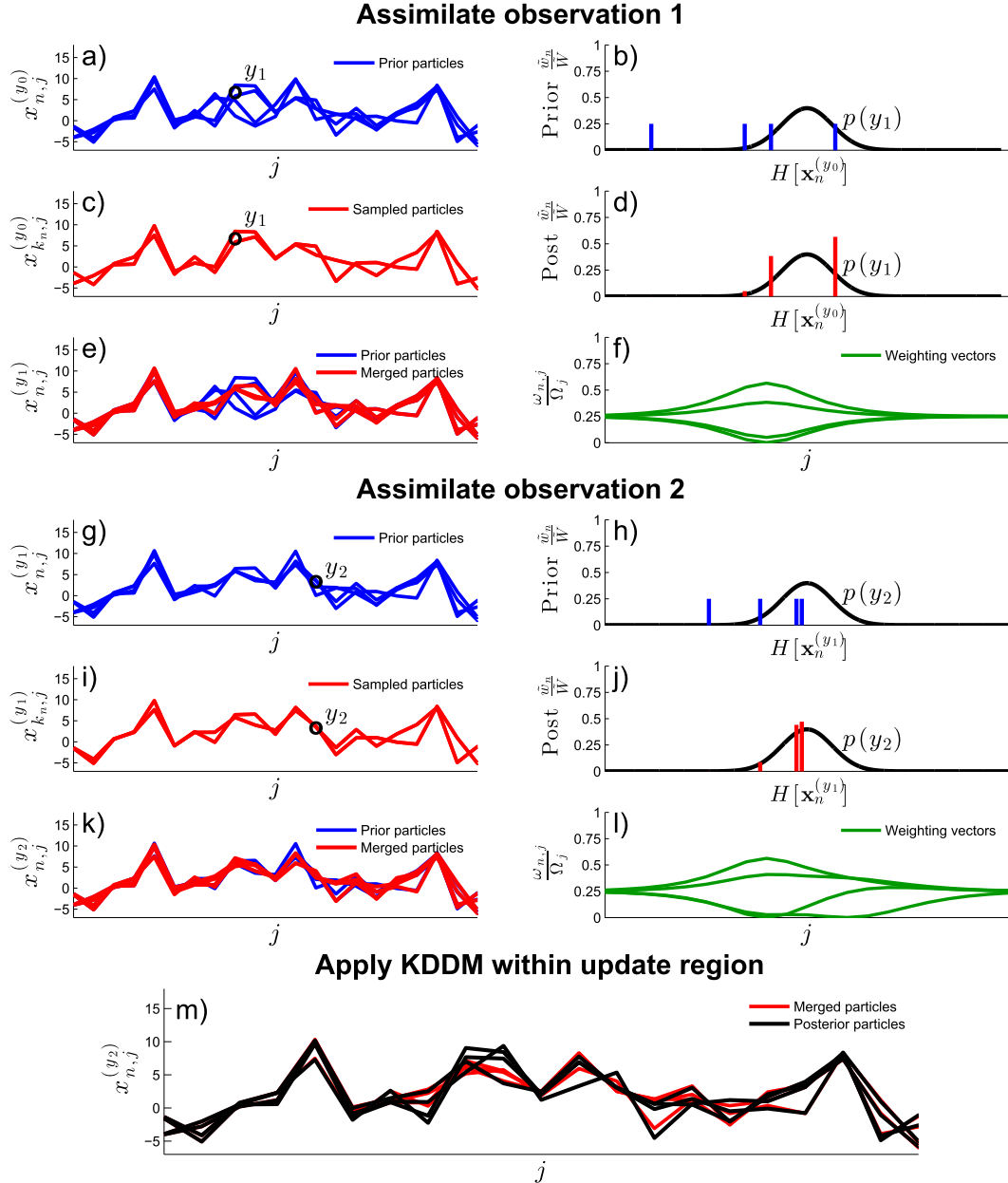
Assimilate observation 1

Assimilate observation 2

Apply KDDM within update region

FIG. 4. Schematic illustrating how the local PF assimilates a pair of observations for the L96 model.

$$r_{1,j} \leftarrow \sqrt{\dfrac{\sigma_j^{(y_i)2}}{\dfrac{1}{N_e-1}\sum_{n=1}^{N_e}[x_{k_n,j}^{(y_{i-1})} - \bar{x}_j^{(y_i)} + c_j(x_{n,j}^{(y_{i-1})} - \bar{x}_j^{(y_i)})]^2}}$$

$$r_{2,j} \leftarrow c_j r_{1,j}$$

$$x_{n,j}^{(y_i)} \leftarrow \bar{x}_j^{(y_i)} + r_{1,j}(x_{k_n,j}^{(y_{i-1})} - \bar{x}_j^{(y_i)}) + r_{2,j}(x_{n,j}^{(y_{i-1})} - \bar{x}_j^{(y_i)})$$

end for

end for

Apply probability mapping for higher-order corrections to posterior sample.

4. Cycling data assimilation experiments

a. Test problem: The Lorenz (1996) model

In this section, the local PF is applied for the L96 model to test the localization and update strategies, and compare the new method with the EAKF. The L96 model contains N_x equally spaced variables, x_i for $i = 1, \ldots N_x$, which are evolved in time using the set of differential equations:

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F, \qquad (19)$$

with cyclic boundaries: $x_{i+N_x} = x_i$ and $x_{i-N_x} = x_i$. The three terms in (19) are analogous to advection, damping, and forcing terms found in geophysical models, and the system exhibits varying degrees of chaotic behavior depending on the choice of F and N_x. For experiments performed in this study, N_x and F remain fixed at 40 and 8, respectively, which leads to chaotic behavior in the system dynamics (Lorenz 1996; Lorenz and Emanuel 1998). Forward integration of (19) is performed numerically using the fourth-order Runge–Kutta method with a time step of 0.05 time units [defined arbitrarily as 6 h; see Lorenz (1996)]. The L96 model and EAKF data assimilation system used in this study are included in the open source NCAR DART software package (Anderson et al. 2009; available online at http://www.image.ucar.edu/DAReS/DART.

b. Experiment setup

Several data assimilation experiments are constructed to compare the PF and EAKF for model and observing-system configurations that mimic applications found in geoscience. These experiments also provide an opportunity to examine the sensitivity of the filters to ensemble size and localization over a variety of problems. Both the PF and EAKF use a Gaspari and Cohn (1999) correlation function for localization with a shape that is modified by specifying a half-width r for its decay to 0.[5] In addition to localization, the PF and EAKF each contain a secondary mechanism for preventing filter divergence during cycling. The EAKF uses the Anderson (2007) adaptive state-space inflation scheme with the prior standard deviation for the inflation coefficient γ fixed at 0.1. This choice of γ is consistent with previous studies using the EAKF with adaptive inflation for the L96 model (e.g., Anderson 2007, 2009). For the PF, the localization function is multiplied by the coefficient α as described in section 3. The optimal α for cases examined in this study ranges from 0.95 to 1, depending on ensemble size.

The relative performance of the PF with respect to the EAKF is examined for a number of model and filter configurations. These experiments include three forms of measurement operator H and several different system parameters, such as the time between observations (Δt) and the number of observations. Unless stated otherwise, the default configuration for these experiments consists of a linear interpolation for H, $\Delta t = 6$ h, $N_y = 20$, and no model error. Observations are generated from a "truth" simulation using (1) with $\epsilon \sim N(0, \mathbf{I})$. Spatial observation locations are fixed with time and chosen

randomly from a Gaussian distribution centered on variable 20 with a standard deviation of $\frac{1}{5}$ the domain length. The resulting network has most of the observations in one region of the model domain, leaving portions of the domain relatively unobserved in a manner similar to environmental observing platforms. The observations are assimilated over 10 000 cycles to provide a large sample for verifying the performance of the two data assimilation systems. After a 1000-cycle spinup period, domain-averaged prior RMSE and ensemble spread are averaged over the remaining 9000 cycles to summarize results from the experiments. The PF and EAKF are compared in this section for optimally tuned values of localization (and α for the PF). The "optimal" configuration of the PF and EAKF are found from off-line sensitivity experiments using ranges of r and α to find the configuration that yields the lowest prior RMSEs. These values include r every 0.01 units from 0.02 to 0.10, every 0.03 units from 0.12 to 0.24, and every 0.10 units from 0.30 to 0.50. Likewise, the local PF is tested for values of α between 0.9 and 1.0 every 0.01 units. The optimal system parameters found in these tests will be discussed at the end of the section.

c. Sensitivity to ensemble size and observation type

For the first test of the local PF, both filters assimilate observations using ensemble sizes of 5, 10, 20, 40, 100, 200, 500, and 1000 particles. Three separate sets of experiments are performed, each differing only in the specification of H: the first experiment uses an interpolation from model space to observation space for H, the second experiment extends H to include $|x|$, and the third experiment applies $\ln(|x|)$ to the interpolated values. Given a univariate random variable $x \sim N(1, 1)$, Fig. 5 illustrates the effects of transforming Gaussian samples into observation space for each $H[x]$. For reference, the red dashed lines indicate the Gaussian estimate of the observation-space priors, calculated from the mean and variance of the transformed sample. The two nonlinear measurement operators introduce an additional source of non-Gaussianity that may limit the effectiveness of the EAKF.

Figure 6 shows average prior ensemble mean RMSEs and spread for each observation type and N_e using the three versions of H presented in Fig. 5. The local PF is applied without KDDM (black lines) and with KDDM (blue lines) to compare the benefits of using probability mapping when updating posterior particles. KDDM requires about 40 particles to provide stable filtering results, owing to the kernel approximation of pdfs in the algorithm; for this reason, Fig. 6 does not show results using KDDM with small ensembles. Experiments using nonlinear H show marginal improvements in the PF's

---

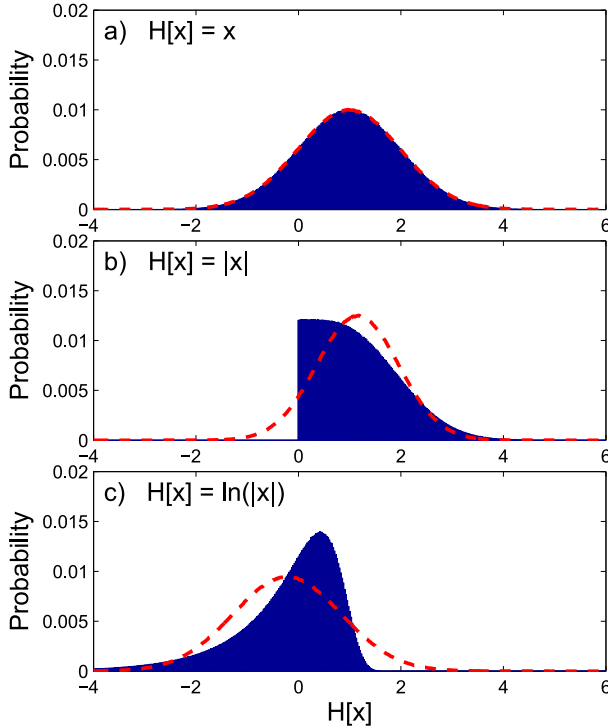[5] Values of r represent the fraction of the L96 model domain length.

FIG. 5. Probability densities demonstrating the non-Gaussian effects of three different measurement operators. The red dashed line shows a Gaussian curve using the same mean and variance as the probability densities given by the blue shading.

accuracy when probability mapping is used, thus demonstrating the performance benefits of applying this method. For future filter comparisons in this study, the KDDM step will be used whenever $N_e \geq 40$.

In general, the PF provides satisfactory results for N_e as small as 5; that is, the variance in the ensemble matches the true forecast error on average, and the prior errors remain lower than the climatological errors for this system—which are found to be about 4.1 from long free-running forecasts. Because these results are unattainable from standard PF approaches, the experiments confirm that localization accomplishes the goal of preventing filter degeneracy for small ensembles.

When H is linear, the PF requires 200 members to match the performance of the EAKF (Fig. 6a). The relative performance of the PF with respect to the EAKF depends on the underlying error distributions of prior quantities and the ensemble size used for approximating probabilities. The PF provides an advantage over the EAKF only when ensemble forecasts present enough evidence to suggest the true prior error distribution is non-Gaussian with some degree of confidence. To determine whether significant differences from normality exist in prior pdfs, the Kolmogorov–Smirnov (KS) test is applied to prior ensembles from PF data
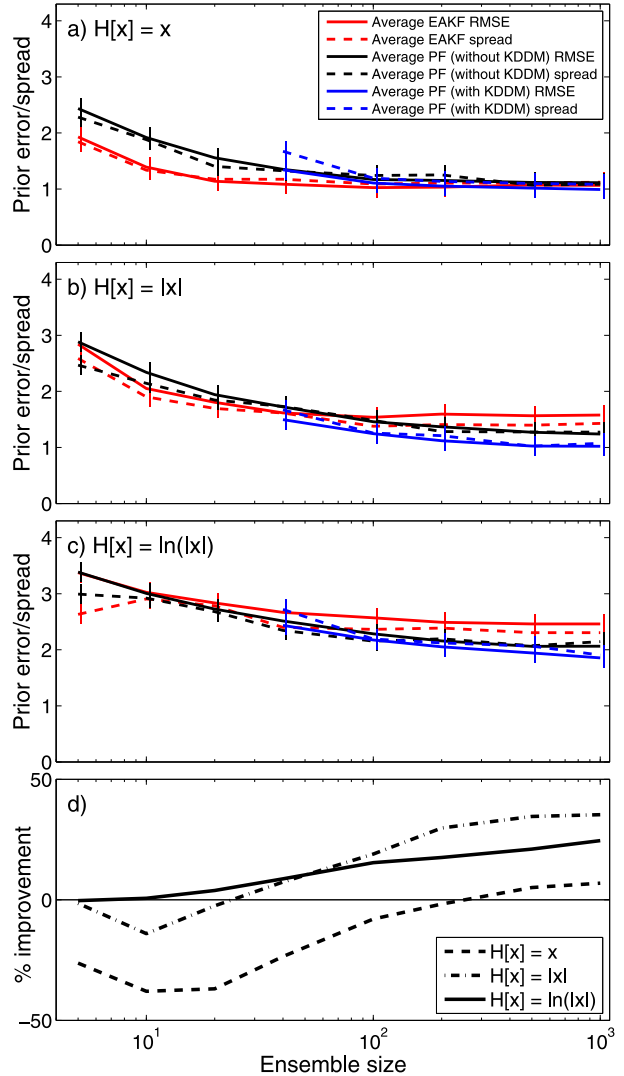
FIG. 6. Prior mean RMSE and spread as a function of ensemble size using (a)–(c) three different types of observations, and (d) percent improvement of PF over EAKF for each experiment. Tick marks on curves in (a)–(c) indicate data points.

assimilation cycles. Table 1 shows the percentage of cycles in which the prior sample for the fifth state variable[6] fails the KS test at the 5% significance level, which provides an estimate of how frequently the ensemble can detect deviations from normality. The first column of the table contains percentages from the linear H experiment for each ensemble size tested. Because of sampling uncertainty, values remain below 10% until N_e reaches about 200, which agrees well with the N_e

---

[6] Variable 5 is chosen for this test because it falls between the densely observed and sparsely observed regions of the model domain.

TABLE 1. Percentage of cycles containing a prior sample for variable 5 that fails the Kolmogorov–Smirnov test at the 5% significance level, tabulated for a range of system parameters. Unless specified otherwise, results are from experiments using the default filter and observation network configuration.

N_e	% cycles	$H[x]$	% cycles	N_y	% cycles	Δt (h)	% cycles
10	0.00	x	7.44	10	0.15	3	6.20
20	0.04	$|x|$	10.51	20	7.44	6	7.54
40	0.54	$\ln(|x|)$	51.62	40	6.58	12	11.63
100	7.44	—	—	100	2.68	24	8.22
200	25.33	—	—	—	—	48	4.01
500	49.68	—	—	—	—	—	—
1000	69.10	—	—	—	—	—	—

required for the PF to begin producing smaller RMSEs than the EAKF (Fig. 6a).

When H is nonlinear, the number of particles required for the PF to provide more accurate solutions than the EAKF decreases. For example, the experiment presented in Fig. 6b introduces nonlinearity by observing the absolute value of the system state. This observation type poses problems for Gaussian filters when the prior ensemble contains both positive and negative values following the interpolation of the model state to observation locations. Because the EAKF is suboptimal for this observation type, its performance is surpassed by the PF for ensembles containing fewer than 40 particles. When $H = \ln|\mathbf{x}|$, the additional source of nonlinearity

allows the PF to outperform the EAKF using as few as 10 particles (Fig. 6c). For this case, the PF requires only five particles to provide forecast results that are as accurate as the EAKF. The second column of Table 1 contains the percentage of times the observation-space priors deviate from normal for $N_e = 100$. The KS test shows increasingly higher occurrences of non-Gaussian probabilities from the three operators tested in this study, thus explaining why the PF becomes progressively more beneficial in each filtering experiment.

To provide a more complete assessment of filter performance when non-Gaussian priors are detected frequently, Fig. 7 shows rank histograms calculated from PF and EAKF priors during the two nonlinear H
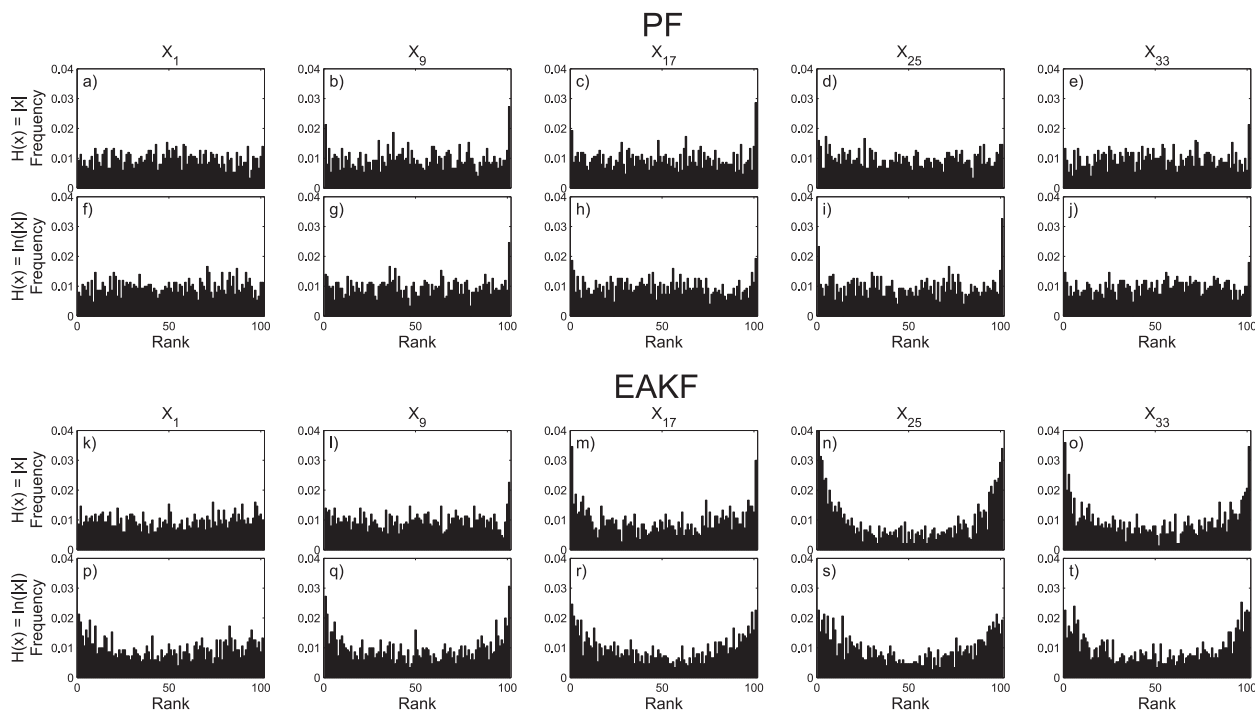

FIG. 7. Rank histograms calculated from (a)–(j) PF and (k)–(t) EAKF priors for the $H[x] = |x|$ and $H[x] = \ln(|x|)$ experiments using $N_e = 100$. (from left to right) The verification is performed for state variables 1, 9, 17, 25, and 33.

experiments. Histograms are calculated for every eighth state variable in the model domain using the $N_e = 100$ cases. This verification counts the number of times the true model state lands in discrete bins formed by ranking prior particles in ascending order (Anderson 1996; Hamill and Colucci 1996; Talagrand et al. 1997). Non-uniform distributions of frequencies across the bins can indicate deficiencies in probabilistic forecasts; for example, the EAKF experiments demonstrate clear deviations from the expected flat distribution (e.g., Figs. 7m–o,r–t). On the other hand, rank histograms produced from PF forecasts are relatively uniform, with the exception of a few variables that yield a larger than ideal number of occurrences where the truth lies outside the range of particles (e.g., Figs. 7b,c,i).

Though not shown, the perturbed-observation EnKF (Houtekamer and Mitchell 1998) is also applied for experiments with nonlinear H. Lei et al. (2010) found this method to be more stable than deterministic filters, such as the EAKF, when prior ensembles exhibit strong non-Gaussianity. Nevertheless, simulations performed with the perturbed-observation filter yielded no significant benefits over the EAKF for the tested applications—possibly due to the use of localization and adaptive inflation to treat systematic errors during data assimilation.

d. Sensitivity to observation network

In this subsection, the filters are applied for various observation networks to explore potential deficiencies in the local PF algorithm that may arise for applications with sparse and dense observations. These tests are performed using N_e fixed at 100 and a linear H. With this configuration and the default observation network, the PF performs almost as well as the EAKF (see Fig. 6a), providing a natural choice of parameters for this comparison. The 100-member ensembles are also small enough to allow for many simulations to be performed at a low computational cost. Using the approach described at the beginning of this section, observation networks are generated by choosing random observation locations for $N_y = 10, 20, 40,$ and 100.

Figure 8a compares prior ensemble mean RMSEs and spread from the cycling experiments after tuning the data assimilation parameters for each observation network. When N_y is decreased to 10, the resulting increase in prior ensemble spread leads to a smaller number of particles landing in the high likelihood region where observations are present. Nevertheless, errors in the PF solution increase for smaller N_y at nearly the same rate as the EAKF, suggesting that localization continues to maintain the filter's stability when fewer observations are available. Large N_y poses additional challenges for PFs, because dense observation networks are more
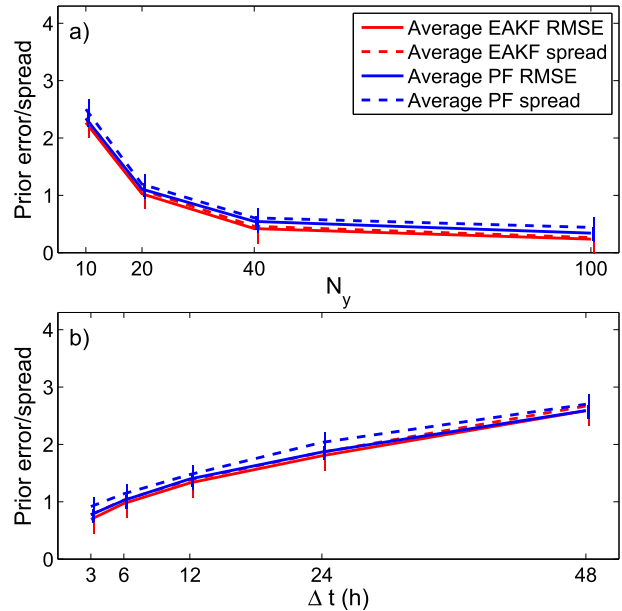

FIG. 8. Prior mean RMSE and spread as a function of (a) observation density and (b) observation frequency for the EAKF and PF. Tick marks on each curve indicate data points.

likely to capture a higher number of degrees of freedom in the system than sparse networks; this assumption is implicit in studies analyzing the asymptotic behavior of PFs (see section 2b). Despite this drawback, the local PF continues to provide accurate results as N_y is increased for a constant N_e.

These experiments demonstrate that the PF and EAKF respond similarly to changes in the spatial density of observations. Likewise, the PF does not produce more accurate results than the EAKF during these experiments. This result occurs because none of the observation networks tested in this section increase the occurrence of non-Gaussian priors during the cycling data assimilation (see third column of Table 1). Though not shown, additional members are required for the PF to outperform Gaussian methods for the tested observation networks, because the EAKF solution is quasi-optimal.

The performance of the PF is also examined for observation networks that measure the state at different frequencies. Here, Δt values of 3, 6, 12, 24, and 48 h are used, with N_y fixed at 20. Because the filtering steps in this experiment occur at different time intervals, results are verified every 48 h to coincide with the largest Δt. The number of verified cycles, however, is kept consistent with previous 6-h cycling experiments by performing the data assimilation over 8 times as many days. Similar to the results using a range of N_y, simulations show that the local PF exhibits sensitivity to temporal observation density that is consistent with the EAKF
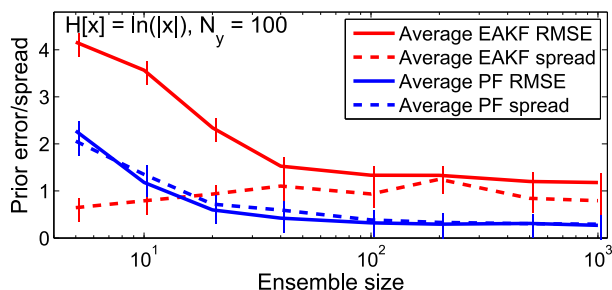
FIG. 9. Prior mean RMSE and spread as a function of ensemble size $H[x] = \ln(|x|)$. Tick marks on each curve indicate data points.

FIG. 10. Optimal localization half-width as a function of ensemble size for each experiment. Tick marks on each curve indicate data points.

(Fig. 8b). Furthermore, the PF provides no practical benefit over the EAKF when the L96 system state is measured infrequently, which follows from the decreasing percentage of non-Gaussian priors for $\Delta t >$ 12 h (fourth column of Table 1). Increasing Δt in these experiments allows forecast errors to approach the climatological errors for the L96 system, which are approximately Gaussian. Though not included in Table 1, additional experiments using 1000 particles to assimilate observations every 48 h yield a similar percentage of cycles that deviate from normal as the 6-h cycling case (71.68% vs 69.10%).

e. Practical advantages of the local PF

Experiments presented in previous parts of this section demonstrate the feasibility of a localized PF for data assimilation through systematic testing of the filter over a range of configurations. This subsection highlights some of the practical advantages of the local PF for real environmental analysis and prediction problems.

For the modeling system tested in this study, the KS test results in Table 1 suggest that the PF provides the largest benefits when N_e is large or when H is nonlinear—otherwise, non-Gaussian priors are detected infrequently. The latter of the two cases is a common occurrence for filtering problems in geophysics. For example, an application relevant for weather analysis and forecasting is the assimilation of remotely sensed data from satellites and radars. Here, filters must process densely spaced observations that relate nonlinearly to the model state, which challenges the Gaussian assumptions of Kalman filtering–based methods. An analog to this problem is tested for the L96 system by assimilating a network of 100 randomly located observations every 6 h using $H[x] = \ln(|x|)$. Results are presented in Fig. 9 for ensemble sizes ranging from 5 to 1000. For this configuration, the PF provides substantial benefits over the EAKF (even for $N_e = 5$), owing to the large number of observations and nonlinearity in the measurement operator. This example
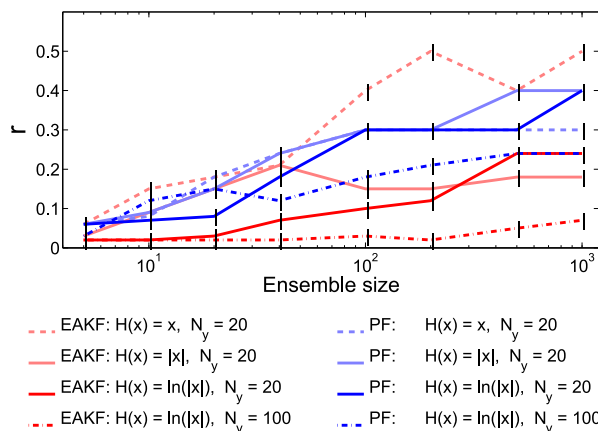
presents a case where the PF can extract information from a dense observation network much more effectively than the EAKF.

Another benefit of the PF over the EAKF is that the optimal localization half-width is less sensitive to the type of observation being assimilated. Figure 10 shows optimal r as a function of N_e for the three measurement operators used in Fig. 6, and the example presented in Fig. 9. The main objective of localization is to reduce sampling errors resulting from the approximation of prior pdfs with finite-sized ensembles of imperfect model forecasts. Nevertheless, the r required to prevent filter divergence in EAKF experiments with nonlinear H (section 4a) is found to be much smaller compared to experiments with linear H. Two possible explanations for this results are the following: 1) the EAKF uses localization to cope with systematic errors—in addition to sampling error—that may occur during the data assimilation, such as incorrect assumptions regarding the linearity of H; or 2) suboptimal estimates of posterior particles lead to errors in succeeding forecasts, which introduces additional sampling errors in prior ensembles. Since the PF makes no assumptions about H, the optimal r in these experiments depends mostly on N_e alone. Though not shown, the optimal r from observation frequency and density tests in section 4d also exhibits a larger sensitivity in EAKF simulations than in PF simulations. The r used to produce the EAKF results in Fig. 8 ranges from 0.21 to 0.50, while the optimal r in PF cases remains close to 0.30, regardless of the observation network.

5. Conclusions

This paper introduces a localized PF for high-dimensional nonlinear filtering applications. The new filter calculates a

vector of posterior weights for each particle based on the likelihood of observations within neighboring regions of each model state variable. Similar to localized EnKFs, the local PF reduces the influence of distant observations on posterior weights by exploiting the fact that the signal-to-noise ratio of cross-variable covariances tends to decrease at large distances in geophysical systems. Because the localized weights depend on much fewer observations than what is used by traditional PFs, the local PF does not require an ensemble size that increases exponentially with the dimension of the system. To generate samples from the posterior density, the local PF processes observations serially, while sequentially updating each particle. The first step in this process is to sample particles with replacement based on the likelihood of the current observation given each particle. Particles removed during sampling are replaced with linear combinations of sampled particles and prior particles in a manner that satisfies the posterior mean and variance, and cross-variable correlations are updated via the sampling part of the algorithm. The sample is then modified to match higher-order statistics of the posterior by applying a nonparametric probability mapping method. These steps result in an ensemble of unique model states that reflect the marginal Bayesian posterior density near observations and the prior density away from observations. Therefore, the filter approximates the standard particle filter solution when the number of particles is too small to prevent filter degeneracy—similar to localization in EnKFs.

In addition to avoiding the assumptions of EnKFs, the new filter is designed to be computationally affordable for large applications. For each observation, the local PF algorithm requires updating the weighting vectors, resampling particles, and calculating the coefficients needed for the update equation; these calculations are made within the localization region only and are easily parallelized. As a result, the cost of performing the particle update, before probability mapping, is comparable to the EAKF data assimilation system in DART. Applying the probability mapping step nearly doubles the computing time needed for the local PF in the experiments performed in this study, but the relative cost of this method becomes trivial as the number of observations increases. Another practical advantage of the local PF is that it does not require stochastic forcing terms in the model dynamics, as is often the case for PF methods that rely on a proposal density to prevent filter degeneracy. Because errors in atmospheric and oceanic models are often unknown, ensembles are typically evolved in time with a deterministic model.

The local PF algorithm presented in this manuscript has qualities that may be problematic for certain geophysical applications. Like EnKFs, the localized updates are not guaranteed to preserve physical balances in the posterior model state. Mitchell et al. (2002) show that imbalance issues are made worse for EnKFs as the localization length scale decreases, which is equally true in a PF framework. Preliminary results assimilating observations in an atmospheric model (not shown) suggest that particles exhibit similar levels of imbalance when updated using the local PF and EAKF. The local PF also assumes observation errors are uncorrelated in order to localize the posterior update provided from each observation; the same assumption is made in sequential EnKF algorithms, such as the EAKF used here. This assumption may not be valid for certain remotely sensed observations, such as satellite retrievals (Stewart et al. 2008). Nevertheless, both issues exist for most data assimilation methods used regularly for atmospheric and oceanic models. Furthermore, the filter may still need large numbers of particles for applications where the degrees of freedom in the system cannot be isolated easily using localization.

The local PF has been added to the NCAR DART software package for thorough testing with the 40-variable Lorenz system. Results from 10 000-cycle data assimilation experiments show that the new filter requires only 5 particles to prevent filter degeneracy for this model. With linear measurement operators and approximately Gaussian priors, the local PF produces lower prior mean RMSEs than the DART EAKF (with localization and adaptive inflation) for ensemble sizes larger than 200. The largest benefit of the local PF occurs in applications where dense networks of observations that relate nonlinearly to the model state are assimilated. In this case, the local PF provides substantial benefits over the EAKF using as few as five particles. These results provide an incentive to explore the potential of the local PF for high-dimensional geophysical systems, such as weather and ocean models. Possible applications include the assimilation of remotely sensed data such as satellite radiances or radar reflectivity, which require highly nonlinear measurement operators. The author has already begun exploring the feasibility of this method in larger models. Though not discussed in the current study, the local PF requires only 25 members to estimate accurate posterior statistics in cycling data assimilation experiments performed with a coarse-resolution atmospheric general circulation model containing 28 200 variables. Data assimilation tests with this model will be the topic of a future study examining the limitations

and benefits of the new filter for high-dimensional problems.

APPENDIX

Derivation of Update Equations

This appendix describes how the coefficients used in the PF update equations are formed. As discussed in section 3, the derivation begins by assuming that the jth state variable of the nth particle is updated according to

$$x_{n,j}^{(y_i)} = \bar{x}_j^{(y_i)} + r_{1,j}(x_{k_n,j}^{(y_{i-1})} - \bar{x}_j^{(y_i)}) + r_{2,j}(x_{n,j}^{(y_{i-1})} - \bar{x}_j^{(y_i)}). \quad (A1)$$

The sample mean of the updated particles in (A1) is first set equal to the posterior mean:

$$
\begin{aligned}
\bar{x}_j^{(y_i)} &= \frac{1}{N_e} \sum_{n=1}^{N_e} [\bar{x}_j^{(y_i)} + r_{1,j}(x_{k_n,j}^{(y_{i-1})} - \bar{x}_j^{(y_i)}) + r_{2,j}(x_{n,j}^{(y_{i-1})} - \bar{x}_j^{(y_i)})], \\
&= \bar{x}_{n,j}^{(y_i)} + r_{1,j}\left(\frac{1}{N_e} \sum_{n=1}^{N_e} x_{k_n,j}^{(y_{i-1})} - \bar{x}_j^{(y_i)}\right) + r_{2,j}(\bar{x}_j^{(y_{i-1})} - \bar{x}_j^{(y_i)}), \\
&\approx \bar{x}_j^{(y_i)} + r_{1,j}(\tilde{\bar{x}}_j^{(y_i)} - \bar{x}_j^{(y_i)}) + r_{2,j}(\bar{x}_j^{(y_{i-1})} - \bar{x}_j^{(y_i)}).
\end{aligned}
\quad (A2)
$$

The approximation in (A2) comes from replacing the mean of sampled particles with the mean estimated from scalar weights without localization $\tilde{\bar{x}}_j^{(y_i)}$. Solving (A2) for $r_{2,j}$ gives

$$
\begin{aligned}
r_{2,j} &= \frac{\bar{x}_j^{(y_i)} - \tilde{\bar{x}}_j^{(y_i)}}{\bar{x}_j^{(y_{i-1})} - \bar{x}_j^{(y_i)}} r_{1,j}, \\
&= c_j r_{1,j}.
\end{aligned}
\quad (A3)
$$

The sample variance of particles updated by (A1) is also set equal to the posterior variance:

$$
\begin{aligned}
\sigma_j^{(y_i)2} &= \frac{1}{N_e-1} \sum_{n=1}^{N_e} [\bar{x}_j^{(y_i)} + r_{1,j}(x_{k_n,j}^{(y_{i-1})} - \bar{x}_j^{(y_i)}) \\
&\quad + r_{2,j}(x_{n,j}^{(y_{i-1})} - \bar{x}_j^{(y_i)}) - \bar{x}_j^{(y_i)}]^2, \\
&= \frac{r_{1,j}^2}{N_e-1} \sum_{n=1}^{N_e} [x_{k_n,j}^{(y_{i-1})} - \bar{x}_j^{(y_i)} + c_j(x_{n,j}^{(y_{i-1})} - \bar{x}_j^{(y_i)})]^2. \quad (A4)
\end{aligned}
$$

Solving (A4) for $r_{1,j}$ and keeping the positive solution gives

$$r_{1,j} = \sqrt{\frac{\sigma_j^{(y_i)^2}}{\frac{1}{N_e-1} \sum_{n=1}^{N_e} \{x_{k_n,j}^{(y_{i-1})} - \bar{x}_j^{(y_i)} + c_j[x_{n,j}^{(y_{i-1})} - \bar{x}_j^{(y_i)}]\}^2}}. \quad (A5)$$

Using the moment-estimating function (10) to calculate the posterior mean and variance terms, (A3) and (A5) provide expressions for the coefficients $r_{1,j}$ and $r_{2,j}$ needed to update particles using (A1).

Equation (A3) can be simplified to avoid calculating the prior and posterior means required for c_j. The first step is to expand each term in c_j using moment-estimating functions based on the likelihood of y_i, given the current particles:

$$
\begin{aligned}
c_j &= \frac{\bar{x}_j^{(y_i)} - \tilde{\bar{x}}_j^{(y_i)}}{\bar{x}_j^{(y_{i-1})} - \bar{x}_j^{(y_i)}}, \\
&= \frac{\sum_{n=1}^{N_e} \frac{\omega_{n,j}^{(y_i)}}{\Omega_j^{(y_i)}} x_{n,j}^{(y_{i-1})} - \sum_{n=1}^{N_e} \frac{\tilde{w}_n}{\tilde{W}} x_{n,j}^{(y_{i-1})}}{\sum_{n=1}^{N_e} \frac{1}{N_e} x_{n,j}^{(y_{i-1})} - \sum_{n=1}^{N_e} \frac{\omega_{n,j}^{(y_i)}}{\Omega_j^{(y_i)}} x_{n,j}^{(y_{i-1})}}, \\
&= \frac{\sum_{n=1}^{N_e} \left(\frac{\omega_{n,j}^{(y_i)}}{\Omega_j^{(y_i)}} - \frac{\tilde{w}_n}{\tilde{W}}\right) x_{n,j}^{(y_{i-1})}}{\sum_{n=1}^{N_e} \left(\frac{1}{N_e} - \frac{\omega_{n,j}^{(y_i)}}{\Omega_j^{(y_i)}}\right) x_{n,j}^{(y_{i-1})}}.
\end{aligned}
\quad (A6)
$$

The coefficients in (A6) reduce to a constant if $[(\omega_{n,j}^{(y_i)})/(\Omega_j^{(y_i)}) - (\tilde{w}_n)/\tilde{W}][1/(N_e) - (\omega_{n,j}^{(y_i)})/(\Omega_j^{(y_i)})]^{-1}$ is constant for all n. To show that this is the case, the weights are expanded to

$$\omega_{n,j}^{(y_i)} = [p(y_i \mid \mathbf{x}_n^{(y_{i-1})}) - 1]l[x_j, y_i, r] + 1, \quad (A7)$$

$$\Omega_j^{(y_i)} = \sum_{n=1}^{N_e} \omega_{n,j}^{(y_i)}, \quad (A8)$$

$$\tilde{w}_n = [p(y_i \mid \mathbf{x}_n^{(y_{i-1})}) - 1]\alpha + 1, \quad \text{and} \quad (A9)$$

$$\tilde{W} = \sum_{n=1}^{N_e} \tilde{w}_n. \quad (A10)$$

A subtle difference exists between (A7) and the weight equation used throughout the manuscript [see (10)]: (A7) calculates weights based on the current observation and particles that reflect all past observations, while (10) calculates weights based on the current observation, the prior particles before assimilating observations at the current time, and weights that reflect all past observations. The two methods provide equivalent results so long as the updated particles reflect the posterior weights after each observation is processed. After substituting (A7)–(A10) into $[(\omega_{n,j}^{(y_i)})/(\Omega_j^{(y_i)}) - (\tilde{w}_n)/\tilde{W}][1/(N_e) - (\omega_{n,j}^{(y_i)})/(\Omega_j^{(y_i)})]^{-1}$, c_j reduces to an expression that is constant for all n, and depends on the localization function, ensemble size, and sum of likelihoods for the current observation:

$$c_j = \frac{N_e(1 - l[x_j, y_i, r])}{l[x_j, y_i, r]\tilde{W}}. \quad (A11)$$

REFERENCES

Ades, M., and P. J. van Leeuwen, 2015: The equivalent-weights particle filter in a high-dimensional system. *Quart. J. Roy. Meteor. Soc.*, **141**, 484–503, doi:10.1002/qj.2370.

Anderson, J. L., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate*, **9**, 1518–1530, doi:10.1175/1520-0442(1996)009<1518:AMFPAE>2.0.CO;2.

——, 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903, doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.

——, 2007: An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus*, **59A**, 210–224, doi:10.1111/j.1600-0870.2006.00216.x.

——, 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus*, **61A**, 72–83, doi:10.1111/j.1600-0870.2008.00361.x.

——, 2010: A non-Gaussian ensemble filter update for data assimilation. *Mon. Wea. Rev.*, **138**, 4186–4198, doi:10.1175/2010MWR3253.1.

——, T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Avellano, 2009: The Data Assimilation Research Testbed: A community facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296, doi:10.1175/2009BAMS2618.1.

Bengtsson, T., C. Snyder, and D. Nychka, 2003: Toward a nonlinear ensemble filter for high-dimensional systems. *J. Geophys. Res.*, **108**, 8775, doi:10.1029/2002JD002900.

——, P. Bickel, and B. Li, 2008: Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. *Probability and Statistics: Essays in Honor of David A. Freedman*, D. Nolan and T. Speed, Eds., Vol. 2, Institute of Mathematical Statistics, 316–334.

Bickel, P., B. Li, and T. Bengtsson, 2008: Sharp failure rates for the bootstrap particle filter in high dimensions. *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, B. Clarke and S. Ghosal, Eds., Vol. 3, Institute of Mathematical Statistics, 318–329.

Bishop, C. H., and D. Hodyss, 2011: Adaptive ensemble covariance localization in ensemble 4D-VAR state estimation. *Mon. Wea. Rev.*, **139**, 1241–1255, doi:10.1175/2010MWR3403.1.

Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Quart. J. Roy. Meteor. Soc.*, **131**, 1013–1043, doi:10.1256/qj.04.15.

——, P. L. Houtekamer, C. Charette, H. Mitchell, and B. He, 2010: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations. *Mon. Wea. Rev.*, **138**, 1567–1586, doi:10.1175/2009MWR3158.1.

Chorin, A., M. Morzfeld, and X. Tu, 2010: Implicit particle filters for data assimilation. *Commun. Appl. Math. Comput. Sci.*, **5**, 221–240, doi:10.2140/camcos.2010.5.221.

Clayton, A. M., A. C. Lorenc, and D. M. Barker, 2013: Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quart. J. Roy. Meteor. Soc.*, **139**, 1445–1461, doi:10.1002/qj.2054.

Doucet, A., N. de Freitas, and N. Gordon, Eds., 2001: An introduction to sequential Monte Carlo methods. *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, 3–14, doi:10.1007/978-1-4757-3437-9_1.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162, doi:10.1029/94JC00572.

——, and P. J. van Leeuwen, 2000: An ensemble Kalman smoother for nonlinear dynamics. *Mon. Wea. Rev.*, **128**, 1852–1867, doi:10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2.

Frei, M., and H. R. Künsch, 2013: Bridging the ensemble Kalman and particle filters. *Biometrika*, **100**, 781–800, doi:10.1093/biomet/ast020.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757, doi:10.1002/qj.49712555417.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith, 1993: Novel approach to nonlinear/non-Gaussian state estimation. *IEE Proc., F, Radar Signal Process.*, **140**, 107–113, doi:10.1049/ip-f-2.1993.0015.

Haerter, J. O., S. Hagemann, C. Moseley, and C. Piani, 2011: Climate model bias correction and the role of timescales. *Hydrol. Earth Syst. Sci.*, **15**, 1065–1079, doi:10.5194/hess-15-1065-2011.

Hamill, T. M., and S. J. Colucci, 1996: Random and systematic error in NMC's short-range Eta ensembles. Preprints, *13th Conf. on Probability and Statistics in the Atmospheric Sciences*, San Francisco, CA, Amer. Meteor. Soc., 51–56.

——, and C. Snyder, 2000: A hybrid ensemble Kalman filter-3D variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905–2919, doi:10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2.

——, and J. S. Whitaker, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776–2790, doi:10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811, doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.

——, and ——, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137, doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.

Ines, A. V. M., and J. W. Hansen, 2006: Bias correction of daily GCM rainfall for crop simulation studies. *Agric. For. Meteor.*, **138**, 44–53, doi:10.1016/j.agrformet.2006.03.009.

Kuhl, D. D., T. E. Rosmond, C. H. Bishop, J. McLay, and N. L. Baker, 2013: Comparison of hybrid ensemble/4DVar and 4DVar within the NAVDAS-AR data assimilation framework. *Mon. Wea. Rev.*, **141**, 2740–2758, doi:10.1175/MWR-D-12-00182.1.

Lei, J., and P. Bickel, 2011: A moment matching ensemble filter for nonlinear non-Gaussian data assimilation. *Mon. Wea. Rev.*, **139**, 3964–3973, doi:10.1175/2011MWR3553.1.

——, ——, and C. Snyder, 2010: Comparison of ensemble Kalman filters under non-Gaussianity. *Mon. Wea. Rev.*, **138**, 1293–1306, doi:10.1175/2009MWR3133.1.

Liu, C., Q. Xiao, and B. Wang, 2008: An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. *Mon. Wea. Rev.*, **136**, 3363–3373, doi:10.1175/2008MWR2312.1.

Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP: A comparison with 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **129**, 3183–3203, doi:10.1256/qj.02.132.

Lorenz, E. N., 1996: Predictability: A problem partly solved. *Proc. Seminar on Predictability*, Vol. 1, Reading, United Kingdom, ECMWF, 1–18.

——, and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, **55**, 399–414, doi:10.1175/1520-0469(1998)055<0399:OSFSWO>2.0.CO;2.

Majda, A. J., D. Qi, and T. P. Sapsis, 2014: Blended particle filters for large-dimensional chaotic dynamical systems. *Proc. Natl. Acad. Sci. USA*, **111**, 7511–7516, doi:10.1073/pnas.1405675111.

McGinnis, S., D. Nychka, and L. O. Mearns, 2015: A new distribution mapping technique for climate model bias correction. *Machine Learning and Data Mining Approaches to Climate Science*, V. Lakshmanan et al., Eds., Springer, 91–99, doi:10.1007/978-3-319-17220-0_9.

Mitchell, H. L., P. L. Houtekamer, and G. Pellerin, 2002: Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Mon. Wea. Rev.*, **130**, 2791–2808, doi:10.1175/1520-0493(2002)130<2791:ESBAME>2.0.CO;2.

Miyoshi, T., K. Kondo, and T. Imamura, 2014: The 10,240-member ensemble Kalman filtering with an intermediate AGCM. *Geophys. Res. Lett.*, **41**, 5264–5271, doi:10.1002/2014GL060863.

Nakano, S., G. Ueno, and T. Higuchi, 2007: Merging particle filter for sequential data assimilation. *Nonlinear Processes Geophys.*, **14**, 395–408, doi:10.5194/npg-14-395-2007.

Piani, C., J. O. Haerter, and E. Coppola, 2010: Statistical bias correction for daily precipitation in regional climate models over Europe. *Theor. Appl. Climatol.*, **99**, 187–192, doi:10.1007/s00704-009-0134-9.

Pires, C. A., O. Talagrand, and M. Bocquet, 2010: Diagnosis and impacts of non-Gaussianity of innovations in data assimilation. *Physica D*, **239**, 1701–1717, doi:10.1016/j.physd.2010.05.006.

Reich, S., 2013: A nonparametric ensemble transform method for Bayesian inference. *SIAM J. Sci. Comput.*, **35**, A2013–A2024, doi:10.1137/130907367.

Slivinski, L., E. Spiller, A. Apte, and B. Sandstede, 2015: A hybrid particle-ensemble Kalman filter for Lagrangian data assimilation. *Mon. Wea. Rev.*, **143**, 195–211, doi:10.1175/MWR-D-14-00051.1.

Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson, 2008: Obstacles to high-dimensional particle filtering. *Mon. Wea. Rev.*, **136**, 4629–4640, doi:10.1175/2008MWR2529.1.

Stewart, L. M., S. L. Dance, and N. K. Nichols, 2008: Correlated observation errors in data assimilation. *Int. J. Numer. Methods Fluids*, **56**, 1521–1527, doi:10.1002/fld.1636.

Talagrand, O., R. Vautard, and B. Strauss, 1997: Evaluation of probabilistic prediction systems. *Proc. ECMWF Workshop on Predictability*, Reading, United Kingdom, ECMWF, 1–25.

van Leeuwen, P. J., 2003: A variance-minimizing filter for large-scale applications. *Mon. Wea. Rev.*, **131**, 2071–2084, doi:10.1175/1520-0493(2003)131<2071:AVFFLA>2.0.CO;2.

——, 2009: Particle filtering in geophysical systems. *Mon. Wea. Rev.*, **137**, 4089–4114, doi:10.1175/2009MWR2835.1.

——, 2010: Nonlinear data assimilation in geosciences: An extremely efficient particle filter. *Quart. J. Roy. Meteor. Soc.*, **136**, 1991–1999, doi:10.1002/qj.699.

Wang, X., and T. Lei, 2014: GSI-based four-dimensional ensemble-variational (4DEnsVar) data assimilation: Formulation and single-resolution experiments with real data for NCEP Global Forecast System. *Mon. Wea. Rev.*, **142**, 3303–3325, doi:10.1175/MWR-D-13-00303.1.

Whitaker, J. S., and T. M. Hamill, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 3078–3089, doi:10.1175/MWR-D-11-00276.1.

Xiong, X., I. M. Navon, and B. Uzunoglu, 2006: A note on the particle filter with posterior Gaussian resampling. *Tellus*, **58A**, 456–460, doi:10.3402/tellusa.v58i4.14798.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253, doi:10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2.

——, M. Zhang, and J. A. Hansen, 2009: Coupling ensemble Kalman filter with four-dimensional variational data assimilation. *Adv. Atmos. Sci.*, **26**, 1–8, doi:10.1007/s00376-009-0001-8.